



Published in final edited form as:

*Genet Epidemiol.* 2021 July ; 45(5): 471–484. doi:10.1002/gepi.22382.

## Incorporating European GWAS findings improve polygenic risk prediction accuracy of breast cancer among East Asians

Ying Ji<sup>1</sup>, Jirong Long<sup>2</sup>, Sun-Seog Kweon<sup>3,4</sup>, Daehee Kang<sup>5,6,7,8</sup>, Michiaki Kubo<sup>9</sup>, Boyoung Park<sup>10</sup>, Xiao-Ou Shu<sup>2</sup>, Wei Zheng<sup>2</sup>, Ran Tao<sup>11,12,#</sup>, Bingshan Li<sup>1,#</sup>

<sup>1</sup>Vanderbilt Genetics Institute, Department of Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN, USA.

<sup>2</sup>Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA.

<sup>3</sup>Department of Preventive Medicine, Chonnam National University Medical School, Hwasun, Korea.

<sup>4</sup>Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Hwasun, Korea.

<sup>5</sup>Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea.

<sup>6</sup>Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea.

<sup>7</sup>Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, Korea.

<sup>8</sup>Institute of Environmental Medicine, Seoul National University Medical Research Center, Seoul, Korea.

<sup>9</sup>RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

<sup>10</sup>Department of Medicine, Hanyang University College of Medicine, Seoul, Korea.

<sup>11</sup>Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA.

<sup>12</sup>Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA.

### Abstract

Previous GWASs have been largely focused on European (EUR) populations. However, PRSs derived from EUR have been shown to perform worse in non-EURs compared to EURs. In this study, we aim to improve PRS prediction in East Asians (EASs). We introduce a rescaled meta-analysis framework to combine both EUR (N = 122, 175) and EAS (N = 30, 801) GWAS summary statistics. To improve PRS prediction in EASs, we use a scaling factor to up-weight the EAS data, such that the resulting effect size estimates are more relevant to EASs. We then derive PRSs for EAS from the rescaled meta-analysis results of EAS and EUR data. Evaluated in an independent EAS validation dataset, this approach increases the prediction liability-adjusted

#Correspondences should be sent to: Bingshan Li, bingshan.li@vanderbilt.edu or Ran Tao, r.tao@vumc.org.

Conflicts of interest: (None)

Nagelkerke's pseudo  $R^2$  by 40%, 41%, and 5%, respectively, compared to PRSs derived from an EAS GWAS only, EUR GWAS only, and conventional fixed-effects meta-analysis of EAS and EUR data. The PRS derived from the rescaled meta-analysis approach achieved an area under the receiver operating characteristic curve (AUC) of 0.6059, higher than AUC = 0.5782, 0.5809, 0.6008 for EAS, EUR, and conventional meta-analysis of EAS and EUR. We further compare PRSs constructed by SNPs that have different LD scores and minor allele frequencies (MAFs) between EUR and EAS, and observe that lower LD scores or MAF in EAS correspond to poorer PRS performance (AUC=0.5677, 0.5530 respectively) than higher LD scores or MAF (AUC=0.589, 0.5993 respectively). We finally build a PRS stratified by LD score differences in EUR and EAS using rescaled meta-analysis, and obtain an AUC of 0.6096, with improvement over other strategies investigated.

## Keywords

genome-wide association study; polygenic prediction; breast cancer; meta-analysis

---

## Introduction

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death in females worldwide (Torre et al., 2015). Although the rate of getting breast cancer has stabilized in some high-income countries, it continues to rise in most Asian and other low and middle-income countries, stressing the need for establishing early risk prediction and management strategies (Denny et al., 2017). Genetic risk factors play an important role in breast cancer predisposition (Nathanson, Wooster, & Weber, 2001). Large scale genome-wide association studies (GWASs) have identified more than 200 loci to be associated with risk of breast cancer (Cai et al., 2014; Michailidou et al., 2015, 2017; Shu et al., 2020; Zheng et al., 2009, 2013). Polygenic risk score (PRS), a weighted aggregation of risk allele counts derived from GWASs, is emerging as a useful tool for breast cancer risk stratification in multiple populations, including Europeans (EURs) and East Asians (EASs) (Khera et al., 2018; Mavaddat et al., 2019; Wen et al., 2016).

The small sample size in GWASs of non-EUR samples and differences of genetic architecture between EUR and other populations make it challenging to develop powerful and well-calibrated PRSs across diverse populations. To date, large-scale breast cancer GWASs were conducted by the Breast Cancer Association Consortium (BCAC), with summary statistics publicly available for ~123,000 (Michailidou et al., 2015) and ~220,000 EUR samples (Michailidou et al., 2017). The largest non-EUR GWAS was conducted by the Asia Breast Cancer Consortium (ABCC), which included more than 40,000 EAS samples (Cai et al., 2014; Shu et al., 2020; Zheng et al., 2009, 2013). From study of 17 anthropometric and blood-panel traits, applying PRSs derived from EUR GWASs directly to non-EUR samples showed poor transferability in general, with approximately ~37%, ~50%, and ~78% lower prediction  $R^2$  in South Asians, EASs, and Africans, respectively, compared to that in EUR populations (Martin et al., 2019). Therefore, developing polygenic risk prediction models for diverse populations is imperative to translate the GWAS findings to

clinical use. This calls for efforts to improve the performance of PRS in non-EUR samples to mitigate the racial disparity.

Given that non-EUR GWASs are usually of insufficient sample size, and that there is extensive genetic sharing across populations (Barbujani, Magagni, Minch, & Cavalli-Sforza, 1997; Gibbs et al., 2015), there are several recent studies that sought to incorporate a large EUR GWAS and a smaller non-EUR GWAS to improve risk prediction in non-EUR populations. Specifically, Coram et al. proposed a cross-population best linear unbiased prediction method based on multi-component linear mixed models, where SNPs were placed in classes defined by GWAS evidence from different ancestries and allelic effects computed in a population-specific fashion (Coram et al., 2015; Coram, Fang, Candille, Assimes, & Tang, 2017); this method requires individual-level training data in the target population. When only GWAS summary statistics are available, Márquez-Luna et al. constructed a trans-ethnic PRS from a weighted linear combination of PRSs from two populations (Márquez-Luna, Loh, Price, & Price, 2017), which could improve the prediction accuracy for type II diabetes in Hispanic/Latinos and South Asians. Grinde et al. found that this approach did not perform well for several anthropometric, blood count, and blood pressure traits in their Hispanic/Latino cohorts (Grinde et al., 2019). Instead, they proposed to construct PRSs by selecting single-nucleotide polymorphisms (SNPs) and their corresponding weights based on different combinations of EUR GWASs, Hispanic/Latino GWASs, or meta-analyses of both. They found that PRSs using an EUR GWAS for SNP selection and a Hispanic/Latino GWAS or meta-analysis for SNP weights performed the best in their empirical studies. These findings suggested that PRS performances could differ by population- and disease-specific genetic architectures. For breast cancer, EUR-based PRSs were reported to perform equally well in Hispanic/Latinos as in EURs but poorly in African Americans (Allman et al., 2015). The performance of EUR- and EAS-based PRSs in EASs remain unclear. This motivates us to evaluate breast cancer PRS predictions in EASs and develop new strategies to construct PRSs targeting non-EUR populations.

Differential linkage disequilibrium (LD) and minor allele frequency (MAF) are major contributors to the poor transferability of PRSs among populations (Wang et al., 2020). In a PRS of the form  $\sum_j w_j X_j$ , where  $X_j$  and  $w_j$  stand for the standardized genotype and weight for SNP  $j$ , respectively, the ideal weight satisfies  $w_j^2 = h_j^2$ , where  $h_j^2$  is the disease heritability directly contributed by SNP  $j$  (Speed & Balding, 2019). Because the disease heritability contributed by a SNP varies according to local LD (Gazal et al., 2017; Speed, Cai, Johnson, Nejentsev, & Balding, 2017), explicit incorporation of EUR LD information can improve prediction accuracy for EURs (Hu et al., 2017; Marquez-Luna et al., 2019; Vilhjálmsson et al., 2015). In addition, SNPs can serve as better proxies for the underlying “true effects” in populations in which they have high LD scores compared to populations in which they have low LD scores, where the LD score of a SNP is the sum of LD  $r^2$  measured between this SNP and all other SNPs (Bulik-Sullivan et al., 2015). MAF has been used as an indication of the strength of natural selection, thus the differences in strengths of selection between ancestries might have an impact on PRSs (Wang et al., 2020). This motivates us to examine whether modeling the LD and MAF differences between EURs and EASs could improve effect size estimation and genetic risk prediction in EAS populations.

In this paper, we are primarily interested in 1) evaluating the transferability of EUR GWAS data to breast cancer risk prediction in EASs, 2) improving risk prediction for breast cancer in EASs, and 3) exploring effects of LD and MAF differences between EUR and EAS ethnicities in PRS construction. We propose a rescaled meta-analysis framework that upweights EAS signals over EUR signals, yielding effect size estimates closer to the true effect sizes in EASs while taking advantage of the large sample sizes of EUR GWASs. We constructed PRSs using summary statistics from the rescaled meta-analysis of EUR and EAS GWAS data and then evaluated their performances in an independent EAS validation dataset. Our PRS outperforms PRSs derived from the EUR or EAS GWAS alone as well as the conventional meta-analysis of EAS and EUR GWASs. The EUR and EAS GWASs used in the analysis are from the BCAC and ABCC, respectively.

## Results

### Trade-off between training GWAS sample size and matched genetic ancestry in PRS prediction in EASs

We first constructed PRSs using SNP-level effect sizes from single EUR or EAS GWASs and evaluated their performance in an independent EAS validation dataset. To assess the impact of sample size on PRS performance, we used EUR GWAS from the BCAC published in 2015 (62,533 cases and 60,976 controls) (Michailidou et al., 2015) and EAS GWAS from the ABCC (14,958 cases and 15,843 controls) (Cai et al., 2014; Zheng et al., 2009, 2013) for PRS construction; we used an earlier, smaller EUR GWAS rather than the most recent, much larger GWAS (Michailidou et al., 2017) from the BCAC to illustrate our strategy of combining EAS and EUR data so as to avoid the EAS data being overwhelmed by the EUR data. For each training GWAS dataset, we derived predictors based on the “*P*+*T*” method (Materials and methods) and chose parameters that maximize prediction accuracy through cross-validation in an independent EAS validation dataset of 3,853 subjects (1,794 cases and 2,059 controls) (MEGA Shanghai; see Materials and methods and S1 Table).

We reported area under the receiver operating characteristics curve (AUC) of the PRSs along with their *P*-value threshold and number of SNPs included in Fig 1 and Table 1. We found that a single individual sample in the EAS training dataset was substantially more informative about predicting breast cancer risk in the EAS validation dataset compared to that in the EUR training datasets. The best PRS derived from the EAS GWAS yielded an AUC of 0.5782. The best PRS derived from the EUR GWAS, which is ~4 times the sample size of the EAS GWAS, yielded a comparable AUC of 0.5809 in the EAS validation set. These findings are consistent across different *P*-value thresholds used and numbers of SNPs included in the PRSs (Fig 1), demonstrating the trade-off between higher prediction accuracy conferred by the larger EUR sample size and the matched EAS ancestry. Similar findings for type II diabetes in Latinos have been reported before (Márquez-Luna et al., 2017).

To explore whether incorporating GWASs from both the EUR and EAS populations can boost prediction performance, we performed a meta-analysis of the EUR and EAS GWASs and used the resulting summary statistics to derive PRSs. We obtained an AUC of 0.6008 in the validation EAS dataset, which was higher than any PRS derived from the EAS or

EUR GWASs alone. These results demonstrated that combining information from a higher-powered EUR GWAS and ancestry matched EAS GWAS helped improve breast cancer risk predictions in EASs, which was consistent with previous reports for other traits in EASs (Lam et al., 2019) and Hispanic/Latinos (Grinde et al., 2019).

### **Upweighting the EAS effect size estimates in meta-analysis improves PRS prediction in EASs**

In conventional fixed-effect meta-analysis, the effect size of each SNP is calculated by an inverse variance weighted sum of the effect size estimates from the participating GWASs. When the true effect sizes are equal across the GWASs, this formula is optimal and entails no efficiency loss compared to a joint analysis of the GWASs using individual-level data (Lin & Zeng, 2009). However, this formula may not be optimal when the goal is to provide effect size estimate from EAS and EUR GWASs for constructing PRS in EASs. Instead, we propose to tip the trans-ethnic meta-analysis towards the EAS population by up-scaling EAS effect size estimates by a factor of  $\alpha$  in addition to the inverse variance weighting scheme (Materials and methods). This strategy enables us to shrink the meta effect size estimates towards the estimates in the EAS GWAS and increase the power to detect EAS-specific signals. To find a good up-scaling factor  $\alpha$ , we conducted a grid search (i.e.,  $\alpha = 1, 2, 3, 4, 5$ ) and then evaluated the AUC of the resulting PRSs using cross-validation (S1 Fig). The conventional fixed-effect meta-analysis is a special case with  $\alpha = 1$ .

We evaluated these PRSs on the EAS validation dataset and found that up-scaling the EAS GWAS with a factor of two or three in meta-analysis could result in increased predicting accuracy (Fig 2). For example, we obtained an AUC of 0.6059 when using  $\alpha = 3$ , compared to an AUC of 0.6008 in conventional fixed-effect meta-analysis (i.e.,  $\alpha = 1$ ). In general, an up-scaling factor of two or three resulted in better prediction performance than conventional fixed-effect meta-analysis across a range of  $P$ -value thresholds used and numbers of SNPs included in the PRS in our analysis (Fig 2). We conducted one-sided Delong's test and did not observe significant difference between the AUCs of PRSs derived from the rescaled meta-analysis ( $\alpha = 3$ ) and conventional fixed-effect meta-analysis ( $P$ -value = 0.14). This is somewhat expected as recent literature on PRS evaluation also showed small AUC differences between PRSs constructed using different methods using the same training dataset (Khera et al., 2018). We note that the difference between the AUCs of PRSs derived from the rescaled meta-analysis ( $\alpha = 3$ ) and EAS GWAS was statistically significant ( $P$ -value =  $4.6 \times 10^{-4}$ ), so was the difference between the AUCs of PRSs derived from the rescaled meta-analysis ( $\alpha = 3$ ) and EUR GWAS ( $P$ -value =  $1.9 \times 10^{-4}$ ).

### **GWAS effect size heterogeneity is related to LD score differences between ancestries**

Wojcik et al. observed inconsistent effect size estimates between populations (Wojcik et al., 2019), which could contribute to reduced transferability of PRSs between populations. The extent of effect size differences between populations differs across the genome. As LD score measures the tagging capacity of a SNP, a natural topic of investigation is whether the extent of LD differences is related to the extent of effect size differences between EURs and EASs.

To examine this, we used the difference in EAS and EUR LD score (Bulik-Sullivan et al., 2015):  $I_{diff,j} = I_{EAS,j} - I_{EUR,j}$  as an indication of a SNP's tagging capacity divergence between populations, where  $I_{EAS,j}$  and  $I_{EUR,j}$  are the ancestry-specific LD scores in EASs and EURs, respectively, estimated from the corresponding populations in the 1000 Genomes Project (Auton et al., 2015). We partitioned all available SNPs into three equally sized groups: “low EAS/high EUR LD score” group with  $I_{diff}$  less than its 33% quantile, “similar EAS/EUR LD score” group with  $I_{diff}$  between its 33% and 67% quantiles, and “high EAS/low EUR LD score” group with  $I_{diff}$  larger than its 67% quantile. To account for the impact of differential GWAS sample sizes on SNPs' effect size estimates, we calculated “standardized” z-scores by dividing the original z-scores by the square root of GWAS sample size (Wojcik et al., 2019). Then, we compared the standardized z-scores in the EUR and EAS GWASs (denoted as  $z_{EUR}$  and  $z_{EAS}$ , respectively) for SNPs with  $P$ -value  $< 5 \times 10^{-8}$  in either the EUR or EAS GWAS (Fig 3A). In general, we observed a reduction of standardized Z scores in EASs compared to EURs regardless of  $I_{diff}$  categories, with an overall slope of 0.64 ( $z_{EAS} = 0.64 \times z_{EUR}$ , 95% confidence interval: 0.61 - 0.67). In addition, we observed that standardized z-scores tended to be even lower in EASs for SNPs with “low EAS/high EUR LD score” ( $z_{EAS} = 0.52 \times z_{EUR}$ , 95% confidence interval: 0.48 - 0.55), and higher in EASs for SNPs with “high EAS/low EUR LD score” ( $z_{EAS} = 0.83 \times z_{EUR}$ , 95% confidence interval: 0.75 - 0.90), suggesting that LD differences are related to observed effect size differences between populations.

### Effects of LD differences on genetic risk prediction

We further investigated the impact of differential tagging capacity due to differential LD levels between populations on PRS performance in EASs. After classifying SNPs into three groups based on  $I_{diff}$ , we constructed group specific PRSs based on both conventional and rescaled meta-analyses and applied the PRSs to the validation EAS dataset. We observed that the performance of PRS in the low EAS/high EUR LD score group is noticeably lower than that in the other two groups. This is true for both the conventional and rescaled meta-analyses (Fig 4). For example, the AUC of the PRS derived from the conventional meta-analysis using SNPs in the low EAS/high EUR LD score group is 0.5677, while the AUC of the PRSs derived using SNPs in the similar EAS/EUR LD score and high EAS/low EUR LD score groups are 0.5837 and 0.5890, respectively. Comparing the rescaled versus conventional meta-analysis, we found that upweighting the EAS effect size estimates with a factor of two or three resulted in a more dramatic increase in prediction accuracy for SNPs in the low EAS/high EUR LD score and similar EAS/EUR LD score groups, while the performance gain in the high EAS/low EUR LD score group appeared to be minimal.

As the LD difference is likely a key factor contributing to the poor transferability of EUR-derived PRS to non-EUR populations, we also explored whether taking LD differences among population into account has potential to increase PRS accuracy. We constructed an  $I_{diff}$ -stratified PRS by the summation of the three best PRSs within each  $I_{diff}$  group (referred to as “ADD3”). We observed marginal significant improvement of AUC from ADD3 (AUC = 0.6096) over conventional fixed-effect meta-analysis PRS (AUC = 0.6008, one-sided Delong's test  $P$ -value = 0.05).

## Effect of MAF differences on GWAS effect size and genetic risk prediction

As MAF differences between populations might also contribute to GWAS heterogeneity and decreased transferability of PRS, we sought to investigate the effects of MAF empirically. We used a similar approach as in the interrogation of LD differences to partition all available SNPs into three equally sized groups by MAF differences. Specifically, we used the difference in EAS and EUR MAF:  $MAF_{diff,j} = MAF_{EAS,j} - MAF_{EUR,j}$  as an indication of a SNP's MAF divergence between populations, where  $MAF_{EAS,j}$  and  $MAF_{EUR,j}$  are the ancestry-specific MAFs reported by the single population GWASs. We partitioned all available SNPs into three equally sized groups: “low EAS/high EUR MAF” group with  $MAF_{diff,j}$  less than its 33% quantile, “similar EAS/EUR MAF” group with  $MAF_{diff,j}$  between its 33% and 67% quantiles, and “high EAS/low EUR MAF” group with  $MAF_{diff,j}$  larger than its 67% quantile. Then, we compared the standardized z-scores in the EUR and EAS GWASs for SNPs with  $P$ -value  $< 5 \times 10^{-8}$  in either the EUR or EAS GWAS similar to what we did previously for LD differences (Fig 3B). We observed that standardized z-scores tended to be lower in EASs for SNPs with “low EAS/high EUR MAF” ( $z_{EAS} = 0.48 \times z_{EUR}$ , 95% confidence interval: 0.44 - 0.51), and higher in EASs for SNPs with “high EAS/low EUR MAF” ( $z_{EAS} = 0.78 \times z_{EUR}$ , 95% confidence interval: 0.64 - 0.83), suggesting that MAF differences are related to observed effect size differences between populations.

We constructed MAF group-specific PRSs based on both conventional and rescaled meta-analyses and applied the PRSs to the validation EAS dataset. We observed that the performance of PRS in the low EAS/high EUR MAF group is noticeably lower than that in the other two groups (Fig S2). This is true for both the conventional and rescaled meta-analyses. For example, the AUC of the PRS derived from the conventional meta-analysis using SNPs in the low EAS/high EUR MAF group is 0.5530, while the AUC of the PRSs derived using SNPs in the similar EAS/EUR MAF and high EAS/low EUR MAF groups are 0.5876 and 0.5993, respectively.

We explored whether integrating MAF and LD difference information would further improve PRS. We stratified SNPs into nine groups cross-tabulated by the three LD score groups and three MAF groups (Fig S3). We observed that the prediction AUC of PRS derived from SNPs in “low EAS/high EUR MAF + high EAS/low EUR LD score” group was the lowest, while the AUC of PRS derived from SNPs in “high EAS/low EUR MAF + high EAS/low EUR LD score” group was the highest. Then, we evaluated the performance of the added score of the best PRSs from the nine groups (referred to as “ADD9”), similar to what we did with ADD3. We found no improvement on AUC for ADD9 (AUC = 0.6084) compared to ADD3 (AUC = 0.6096), indicating that the impact of MAF differences on PRS prediction might overlap with that of LD score differences (the correlation between MAF differences and LD score differences was 0.36).

## Evaluation of the PRSs using the prediction Nagelkerke's pseudo $R^2$ and odds-ratio in EASs

We evaluated the candidate PRSs using the prediction Nagelkerke's pseudo  $R^2$  and liability-adjusted  $R^2$ . We included the PRSs constructed from the EUR and EAS GWAS alone and the rescaled meta-analyses that upweights the EAS effect size estimates by a factor of two

or three. We also included an  $I_{diff}$ -stratified PRS constructed by the summation of the three best PRSs within each  $I_{diff}$  group. We used logistic regression and included age and the first two principal components as covariates. The results were shown in S2 Table. The PRS derived from the rescaled meta-analysis that upweights the EAS effect size estimates by a factor of three increased the prediction Nagelkerke's pseudo  $R^2$  (liability-adjusted  $R^2$ ) by 40% (41%), 41% (42%), and 5% (6%), respectively, compared to PRSs derived from the EAS GWAS only, EUR GWAS only, and conventional fixed-effects meta-analysis of both. The  $I_{diff}$ -stratified PRS performed better than the other models, although the improvement was marginal (Table 1).

We assessed the odds ratio (OR) of developing breast cancer in the top 10% individuals with the highest PRSs versus the remaining 90% in the EAS validation dataset. We observed that PRSs derived from the meta-analyses of EAS and EUR GWASs resulted in improved ORs compared to PRSs derived from the EAS GWAS or EUR alone (Table 2). For example, we obtained ORs in the range of 2.31 - 2.50 for PRSs derived from the meta-analyses, while we obtained ORs of 2.23 and 1.73 for PRSs derived from the EAS GWAS or EUR, respectively. We also compared the top 10% individuals with the middle 40%-60% and observed similar results (S2 Table).

## Discussion

There has been tremendous progress in discovery of GWAS loci associated with breast cancer, making it feasible to construct PRS for individualized risk stratification. However, there is a lack of well-powered GWAS in non-EUR populations, a challenge that may exacerbate disparity in clinical use. The primary goal of this work is to explore strategies that can improve PRS of non-EUR populations, particularly in EAS. We found that incorporating information from well powered EUR GWAS and explicitly modeling LD and MAF differences are promising to improve PRS for breast cancer risk prediction in EAS. We proposed an approach to construct PRS from a rescaled meta-analysis of EUR and EAS GWAS which upweights the EAS component relative to the conventional inverse-variance based weightings. We observed improvement in PRS prediction accuracy using rescaled meta-analyses. As LD and MAF differences are likely key factors contributing to the poor transferability of EUR-derived PRS to non-EUR populations, we also explored whether taking them into account have potential to increase PRS accuracy. We observed marginal significant improvement of AUC from the  $I_{diff}$ -stratified PRS ("ADD3") over conventional fixed-effect meta-analysis PRS ("META-FE",  $P$ -value = 0.05) but no further improvement by stratifying on both LD and MAF differences ("ADD9").

We further dissect why rescaled meta-analysis strategy is able to increase the PRS accuracy of the risk prediction in EAS in this study and in other non-EUR populations in general. We define the true model as  $y = \beta_{EAS,j}x_j$  for EAS and  $y = \beta_{EUR,j}x_j$  for EUR. Given the genetic differences between EAS and EUR,  $\beta_{EAS,j}$  and  $\beta_{EUR,j}$  are often unequal, and the extent of their differences depends on many factors, including LD, MAF, and environment factors. To construct a powerful PRS for EAS, i.e.,  $w_{EAS,j}x_j$  the goal is to assign a weight for the  $j$ th SNP  $w_{EAS,j}$  that is close to the true effect size  $\beta_{EAS,j}$ . When we use weights derived from EAS GWAS as  $w_{EAS,j}$  the estimates are unbiased but of larger variance due



to the smaller sample size. On the other hand, when the weights are derived from EUR GWAS, the estimates have smaller variance but are biased away from the true effect sizes in EAS, leading to poor transferability of EUR-derived PRS to EAS populations. To obtain estimates with a better bias and variance tradeoff, we proposed to combine EAS and EUR data to obtain estimates for use in PRS construction in EAS. Specifically, we proposed a rescaled meta-analysis strategy, with a rescale factor  $\alpha > 1$  to “pull” the estimate towards the true effect size in EAS, i.e.  $w_{EASj} = \alpha * w_{inv\_EAS,j} \beta_{EASj} + w_{inv\_EUR,j} \beta_{EURj}$ , where  $w_{inv\_EAS,j}$  and  $w_{inv\_EUR,j}$  are the inverse-variance weights in the conventional meta-analysis (Materials and methods). The magnitude of  $\alpha$  controls for the extent of “pulling” towards EAS, with  $\alpha = 0$  or  $\alpha = inf$  corresponding to the extreme case where only EUR data ( $\alpha = 0$ ) or only EAS data ( $\alpha = inf$ ) are used. For  $\alpha > 1$ , it upweights EAS in a way that a sample in the EAS GWAS data contributes more than a sample in EUR GWAS to the resulting effect size estimate, thus achieving the effect of pulling toward EAS. The optimal magnitude of  $\alpha$  depends on the relative sample size of EAS and EUR, and in general,  $\alpha$  increases when EAS sample sizes increases relative to EUR due to a bias-variance tradeoff between the more accurate EAS results and more precise EUR results. When the EAS sample size is much smaller than the EUR sample size, the effect size estimates obtained from the EUR GWAS have much smaller standard errors. In this situation, upweighting the EAS effect size with a larger  $\alpha$  may significantly increase the standard errors in the rescaled meta-analyzed compared to that in conventional fixed-effects meta-analysis, offsetting the potential benefit of reduced bias. For example, when we used the latest EUR GWAS data of ~220,000 samples, the optimal  $\alpha$  decreases to 1.3, achieving an AUC = 0.6195, compared to AUC = 0.6119 obtained in the traditional meta-analyses. To comprehensively study the relationship between optimal  $\alpha$  and relative EUR and EAS sample sizes, we would need access to individual-level training data so that we can conduct a series of GWASs at different sample sizes (e.g., 60k, 120k, 180k, and 220k for EUR and 10k, 20k, and 30k for EAS). For a variety of diseases, more non-EUR samples are being generated, e.g., PAGE (Wojcik et al., 2019) and TOPMed (Taliun et al., 2019), and the reweighting factor is expected to increase when meta-analyzing with EUR data, resulting in increased prediction accuracy for non-EUR samples. Note that we used only a single scaling factor  $\alpha$  to controls for the adjustment. Ideally, if we are able to assign a SNP-level scaling factor for every SNP, i.e.  $w_{EASj} = \alpha_j * w_{inv\_EAS,j} \beta_{EASj} + w_{inv\_EUR,j} \beta_{EURj}$  where  $\alpha_j$  is the scaling factor for the  $j$ th SNP, the performance of PRS can be further improved. However, it is challenging to assign scaling factors for each individual risk allele, as it is unknown *a priori* which alleles have different effect sizes between populations, and to what extent of their differences are. It requires further efforts to model fine-scale genetic difference between populations to assign reasonable SNP-level scaling factors.

We also examined the predictive performance on our validation set using Grinde et al.’s approach that performed well in several anthropometric and blood count traits in Hispanic Americans: select SNPs based on European GWASs and use meta-analysis weight estimates to construct PRS. We find resulting PRS perform worse (AUC = 0.5859) than using conventional meta-analysis GWAS for both SNP selection and weight estimates (AUC = 0.6008). This agrees with previous findings that the best performing PRS for a trait likely

depends on the genetic architecture, differences in sample size between populations, and other factors.

We are aware that there are other PRS construction methods besides “ $P+T$ ”, such as LDpred (Vilhjálmsón et al., 2015) and SBayesR (Lloyd-Jones et al., 2019). Specifically, we applied LDpred using the EAS LD reference panel and presented the results in Table S3. The findings were similar compared to those obtained by “ $P+T$ ”: PRS derived from conventional fixed-effects meta-analysis performed better than those derived from single population GWASs; PRSs derived from rescaled meta-analysis and the  $I_{diff}$ -stratified PRS performed better than PRS derived from conventional fixed-effect meta-analysis. We observed comparable performance between PRSs constructed using LDpred and “ $P+T$ ”, which is consistent with existing literature on breast cancer PRS (Khera et al., 2018).

We studied the impact of LD and MAF PRS constructed by taking into account both LD score differences and MAF differences between EAS and EUR did not outperform PRS constructed by taking into account LD score differences alone in EAS. It would be worthy of further exploration on how to better leverage MAF and LD in PRS construction. In addition, we did not explore the impact of functional genomic annotations on trans-ethnic PRS prediction. As previous studies have shown that the use of epigenetic and functional annotations improves heritability estimation and PRS prediction in a single population (Hu et al., 2017; Lloyd-Jones et al., 2019), an interesting topic of investigation is to incorporate those annotations when constructing trans-ethnic PRS to further boosting prediction accuracy.

Our work is based on a target population of EAS, while there are potential opportunities to extend the strategies explored in this study to other ethnicities. For example, the explicit modeling of genetic difference between EUR and African has potential to improve PRS in African and in African Americans. Although our approach has been effective in a relatively homogenous population like EAS, its application remains challenging in admixed populations with complex LD patterns and demographic history like African Americans or Hispanic/Latinos. Since the genomes of admixed individuals are a mosaic of segments with different ancestral origins, a first step would be to get ancestry specific effect size estimates and  $P$ -values from training GWASs, which is often not available from publicly available summary statistics. If individual-level training GWAS data is available, recently developed methods like Tractor (Atkinson et al., 2020) could be applied to obtain ancestry specific summary statistics by generating ancestry dosage at each site from local ancestry inference calls and running a local ancestry-aware regression. Similarly, for the validation data, local ancestry haplotype dosage for each person at each variant need to be estimated and weighted by the ancestry specific effect size estimated in the previous step to allow the generation of “ancestry-specific” PRSs. After that, we can experiment with our strategy of globally upweighting the “more informative” ancestry-level PRS. However, local ancestry estimation in both training and validation sets might introduce bias and the anticipated large sample size discrepancies between EUR and African Americans GWAS studies might further complicate the application. We think this question is worthy of further exploration and we believe that the rapid expansion of genomic resources in admixed populations will be critical to improve PRS predictions. Besides genetic factors (e.g., LD and MAF),

environmental factors also influence effect size differences among ancestries. We argue that, for admixed populations, it is critical for PRS to be ancestry-aware, especially for clinical use, since each individual admixed genome has unique local ancestry profiles, and without taking local ancestry into account it is hard to maintain desired sensitivity and specificity due to genetic differences among ancestries.

In summary, we proposed an approach to construct breast cancer PRS in EAS derived from a rescaled meta-analysis of EUR and EAS GWAS. Different from conventional inverse-variance based weighting framework, our approach upweights the EAS component over the EUR component. PRS derived from our rescaled meta-analysis outperforms PRS derived from single population GWAS or conventional meta-analysis. This strategy of integrating GWASs across ethnicities when building PRS prediction models could potentially be extended to other non-EUR populations.

## Materials and methods

### GWAS training data in samples of European ancestry

Two large, publicly available summary statistics datasets based on European ancestry were used in this study. The data were from the Breast Cancer Association Consortium (BCAC) (details see S1 Table). EUR\_2015 (Michailidou et al., 2015) includes two subsets, GWAS (N = 32,498) and COGS (N = 89,677). EUR\_2017 (Michailidou et al., 2017) is the largest available GWAS study of breast cancer in European ancestry population. This study consisted of three subsets, GWAS, COGS and OncoArray (N = 106,776).

### GWAS training data in samples of Asian ancestry

The GWASs in samples of East Asian ancestry were from the Asian Breast Cancer Consortium (ABCC), which includes 14,958 breast cancer cases and 15,843 controls of Asian ancestry (Cai et al., 2014; Zheng et al., 2009, 2013). Samples were from studies conducted in mainland China, South Korea, Japan, Thailand, Malaysia, Singapore, Canada, U.S., Hong Kong, Taiwan and other countries and regions. Details are in S1 Table. We used the meta-analyzed summary statistics data from the study.

### Validation data of East Asian ancestry

The validation set of EAS ancestry is from the Shanghai breast cancer genetic study, including 1794 cases and 2059 controls. Samples were all genotyped on MEGA chip (Illumina), with ~80k custom Asian content selected to improve the coverage of low-frequency SNPs in Asian populations. Data were imputed using the Phase 3 release of the 1000 Genomes Project as reference.

### Meta-analysis of EUR and EAS

Let  $\beta_{EAS,j}$  and  $V_{EAS,j}$  be the expectation and its variance of SNP  $j$  in EAS GWAS. The corresponding estimates in EUR are  $\beta_{EUR,j}$  and  $V_{EUR,j}$ . In conventional meta-analysis, we have

$$\hat{\beta} = \frac{w_{inv\_EAS,j}\beta_{EAS,j} + w_{inv\_EUR,j}\beta_{EUR,j}}{w_{inv\_EAS,j} + w_{inv\_EUR,j}}$$

where  $w_{inv\_EAS,j} = 1/V_{EAS,j}$  and  $w_{inv\_EUR,j} = 1/V_{EUR,j}$  are the inverse-variance weights for EAS and EUR, respectively. As our goal is to obtain weights to construct PRS for EAS, we multiply a scaling factor (denoted as,  $\alpha > 1$ ) for EAS to the inverse variance weight to obtain a rescaled estimate

$$\hat{\beta} = \frac{\alpha * w_{inv\_EAS,j}\beta_{EAS,j} + w_{inv\_EUR,j}\beta_{EUR,j}}{\alpha * w_{inv\_EAS,j} + w_{inv\_EUR,j}}$$

with variance

$$\hat{v} = \left( \frac{\alpha * w_{inv\_EAS,j}}{\alpha * w_{inv\_EAS,j} + w_{inv\_EUR,j}} \right)^2 V_{EAS,j} + \left( \frac{w_{inv\_EUR,j}}{\alpha * w_{inv\_EAS,j} + w_{inv\_EUR,j}} \right)^2 V_{EUR,j}$$

To select the scaling factor, we tried a grid of  $\alpha = 1, 2, 3, 4, 5$  to derive the resulting summary statistics. Then we selected the scale factor to use along with  $P$ -value threshold that achieved a high prediction accuracy by cross-validation in the EAS validation set.

### “ $P+T$ ”

The  $P+T$  method refers to the calculation of PRS using informed LD-pruning and  $P$ -value thresholding. In this study, we use the implementation of the  $P+T$  method in the software package PRSice-2 (Euesden, Lewis, & O’Reilly, 2015) with the default threshold of  $r^2 = 0.2$  for clumping correlated markers and clumping windows of 250 kb. We examined varying strengths of LD among SNPs by repeating the procedures and changing threshold for clumping correlated markers ( $r^2 = 0.1, 0.2, 0.4, 0.6, 0.8$ ) and the sizes of clumping windows (250 kb, 500 kb), and found results to be similar. For any pair of SNPs that have a physical distance smaller than the clumping window or  $r^2$  greater than the selected threshold, the less significant SNP is removed. PRS is computed by summing risk alleles weighted by effect sizes derived from input summary statistics. The  $P$ -value threshold are selected using validation data to optimize prediction accuracy. We constructed PRS for EAS using selected SNPs and effect size estimates  $PRS = \sum_{j=1}^J w_{EAS,j}x_j$ , which  $w_{EAS,j}$  is the weight for the  $j$ th SNP.

### LDpred

LDpred is a method that uses the GWAS summary statistics and LD information from the external LD reference sample to infer the posterior mean effect size of each SNP, conditioning on the SNP effect estimates of other correlated SNPs. This method assumes a point-normal prior on the distribution of SNP effects such that only a fraction of SNPs have non-zero estimated effects. These fractions of causal SNPs (denoted as  $f$ ) were used in the validation set: 1 (i.e., all SNPs), 0.3, 0.1, 0.03, 0.01, 0.003, and 0.001, with an LD radius of

400 (i.e.,  $M/3000$ ,  $M$  is the number of SNPs, around 1.2 million Hapmap SNPs is included in the current analysis) to obtain local LD information, as suggested by the authors.

### Assessment of PRSs

Our analysis used genotypes and phenotypes in independent validation data of East Asian ancestry from training GWAS. We reported area under the ROC curve (AUC) in a logistic regression model using the disease as outcome. When using model with only PRS as the predictor, we report the in-sample fit using all individuals in the validation set. When using models with PRS, age and first 2 genotype PCs, we use the 10-fold cross-validation procedure. To compare AUCs from different training GWAS data, we conducted one-sided Delong's test for paired AUC curves using "roc.test" implemented in R package pROC.

We also include average Nagelkerke's pseudo  $R^2$  liability-scale pseudo  $R^2$  for the models (Lee, Goddard, Wray, & Visscher, 2012) and the likelihood ratio test  $P$ -value. Nested models are considered to provide performance estimates of PRSs: the full model (PRS + covariates including age and first two PCs of genotype) and the reduced model (covariates only). Nagelkerke's pseudo  $R^2$  was calculated comparing the full model with the reduced model with the covariates alone, thus yielding an estimate of how well the variable (PRS here) explains the data. R packages "rcompanion" (see Web Resources) was used in the analysis. Since Nagelkerke's  $R^2$  suffers from bias when case/control proportion is different, we included liability-scale  $R^2$  that accounted for an ascertained case/control ratio (Lee et al., 2012).

To assess the relationship of PRS with breast cancer case/control status, individuals in the validation set were binned into 10 deciles according to the PRS, and the percentage of cases within each bin was determined. We calculated odds ratio (OR) comparing top 10% of the individuals with the remaining 90% of the samples as reference group, as well as OR comparing top 10% of individuals with individuals in the 40<sup>th</sup>-60<sup>th</sup> percentiles.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgements

This study is supported by NIH grants U01HG009086 (Y.J., R.T., B.L.) and R01HL151152 (R.T.).

---

#### Web Resources

LD score: <https://data.broadinstitute.org/alkesgroup/LDSCORE/>

PRSice2 package (for  $P+T$ ): <https://www.prsice.info/>

Rcompanion package (for pseudo- $R^2$ ): <https://cran.rproject.org/web/packages/rcompanion/index.html>

2015 EUR and 2017 EUR GWAS summary statistics: <http://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/gwas-icogs-and-oncoarray-summary-results/>

Liability-adjusted  $R^2$  implementation: adapted from [https://github.com/armartin/pgc\\_scz\\_asia/blob/master/eur\\_eas\\_prs.R](https://github.com/armartin/pgc_scz_asia/blob/master/eur_eas_prs.R)

## Data Availability Statement

Access to the ABCC data could be requested by submission of an inquiry to Dr. Wei Zheng (wei.zheng@vanderbilt.edu). Request of access to the BCAC data could be submitted directly to BCAC (<http://bcac.ccge.medschl.cam.ac.uk/>).

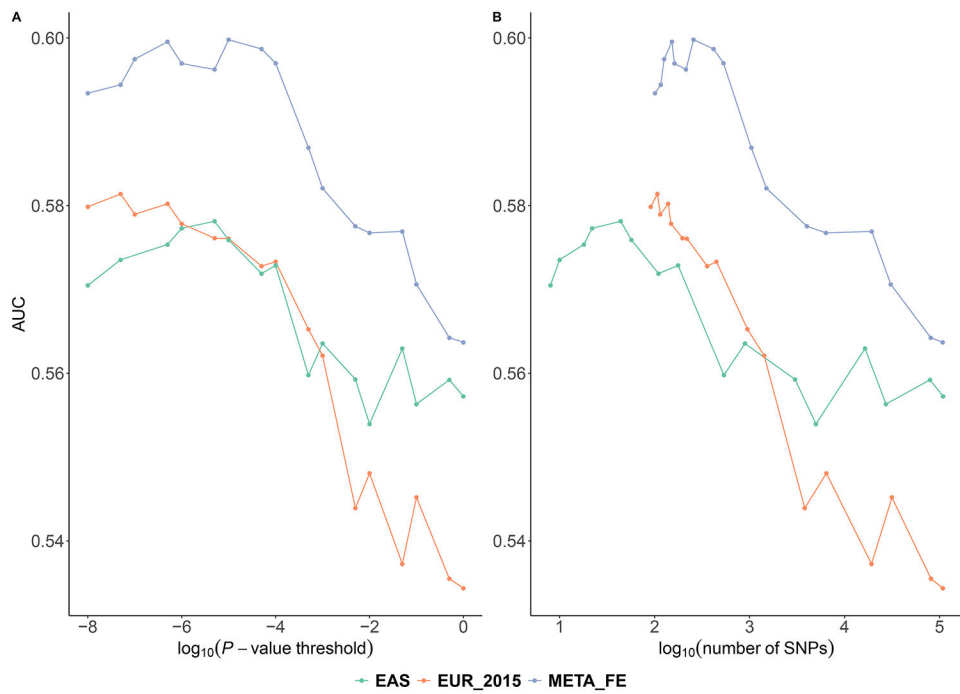
## References

- Allman R, Dite GS, Hopper JL, Gordon O, Starlard-Davenport A, Chlebowski R, & Kooperberg C (2015). SNPs and breast cancer risk prediction for African American and Hispanic women. *Breast Cancer Research and Treatment*, 154(3), 583–589. 10.1007/s10549-015-3641-7 [PubMed: 26589314]
- Atkinson EG, Maihofer AX, Kanai M, Martin AR, Karczewski KJ, Santoro ML, ... Neale BM (2020). Tractor: A framework allowing for improved inclusion of admixed individuals in large-scale association studies. *BioRxiv*, 2020.05.17.100727. 10.1101/2020.05.17.100727
- Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, ... Schloss JA (2015, 9 30). A global reference for human genetic variation. *Nature*. Nature Publishing Group. 10.1038/nature15393
- Bulik-Sullivan B, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, ... O'Donovan MC (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, 47(3), 291–295. 10.1038/ng.3211 [PubMed: 25642630]
- Cai Q, Zhang B, Sung H, Low S-K, Kweon S-S, Lu W, ... Zheng W (2014). Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nature Genetics*, 46(8), 886–890. 10.1038/ng.3041 [PubMed: 25038754]
- Coram MA, Candille SI, Duan Q, Chan KHK, Li Y, Kooperberg C, ... Tang H (2015). Leveraging Multi-ethnic Evidence for Mapping Complex Traits in Minority Populations: An Empirical Bayes Approach. *The American Journal of Human Genetics*, 96(5), 740–752. 10.1016/J.AJHG.2015.03.008 [PubMed: 25892113]
- Coram MA, Fang H, Candille SI, Assimes TL, & Tang H (2017). Leveraging Multi-ethnic Evidence for Risk Assessment of Quantitative Traits in Minority Populations. *The American Journal of Human Genetics*, 101(2), 218–226. 10.1016/J.AJHG.2017.06.015 [PubMed: 28757202]
- Denny L, de Sanjose S, Mutebi M, Anderson BO, Kim J, Jeronimo J, ... Sankaranarayanan R (2017, 2 25). Interventions to close the divide for women with breast and cervical cancer between low-income and middle-income countries and high-income countries. *The Lancet*. Lancet Publishing Group. 10.1016/S0140-6736(16)31795-0
- Euesden J, Lewis CM, & O'Reilly PF (2015). PRSice: Polygenic Risk Score software. *Bioinformatics*, 31(9), 1466–1468. 10.1093/bioinformatics/btu848 [PubMed: 25550326]
- Gazal S, Finucane HK, Furlotte NA, Loh PR, Palamara PF, Liu X, ... Price AL (2017). Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nature Genetics*, 49(10), 1421–1427. 10.1038/ng.3954 [PubMed: 28892061]
- Grinde KE, Qi Q, Thornton TA, Liu S, Shadyab AH, Chan KHK, ... Sofer T (2019). Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. *Genetic Epidemiology*, 43(1), 50–62. 10.1002/gepi.22166 [PubMed: 30368908]
- Hu Y, Lu Q, Powles R, Yao X, Yang C, Fang F, ... Zhao H (2017). Leveraging functional annotations in genetic risk prediction for human complex diseases. *PLOS Computational Biology*, 13(6), e1005589. 10.1371/journal.pcbi.1005589 [PubMed: 28594818]
- Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, ... Kathiresan S (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, 50(9), 1219–1224. 10.1038/s41588-018-0183-z [PubMed: 30104762]
- Lam M, Chen CY, Li Z, Martin AR, Bryois J, Ma X, ... Huang H (2019). Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nature Genetics*, 51(12), 1670–1678. 10.1038/s41588-019-0512-x [PubMed: 31740837]

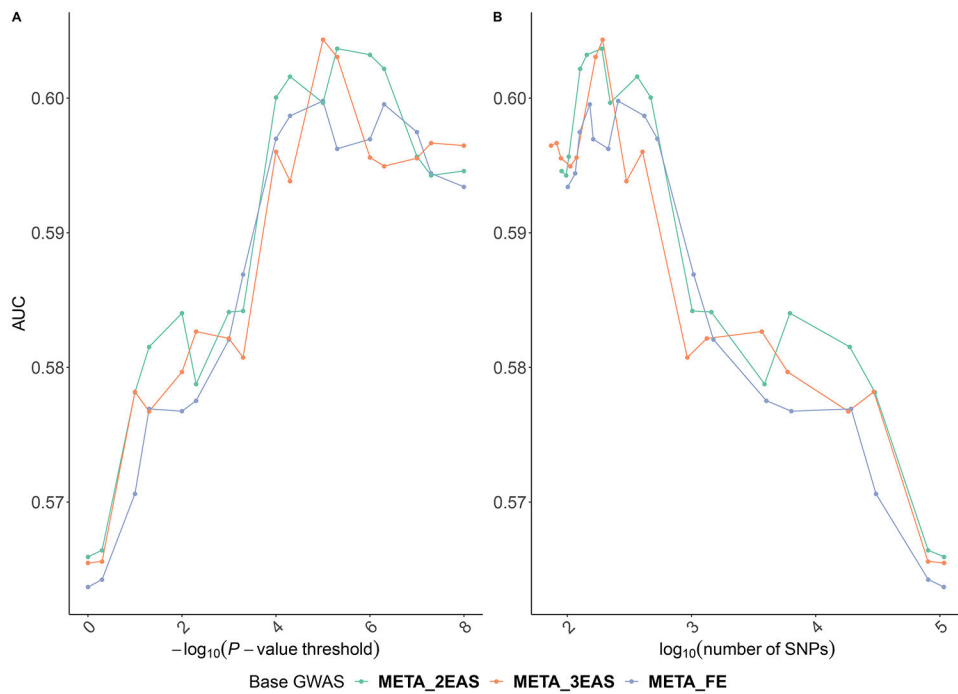
- Lee SH, Goddard ME, Wray NR, & Visscher PM (2012). A Better Coefficient of Determination for Genetic Profile Analysis. *Genetic Epidemiology*, 36(3), 214–224. 10.1002/gepi.21614 [PubMed: 22714935]
- Lin DY, & Zeng D (2009). Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genetic Epidemiology*, n/a-n/a. 10.1002/gepi.20435
- Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, ... Visscher PM (2019). Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nature Communications*, 10(1). 10.1038/s41467-019-12653-0
- Marquez-Luna C, Gazal S, Loh P-R, Furlotte N, Auton A, Price AL, ... Price AL (2019). Modeling functional enrichment improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. *BioRxiv*, 375337. 10.1101/375337
- Márquez-Luna C, Loh P-R, Price AL, & Price AL (2017). Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genetic Epidemiology*, 41(8), 811–823. 10.1002/gepi.22083 [PubMed: 29110330]
- Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, & Daly MJ (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, 51(4), 584–591. 10.1038/s41588-019-0379-x [PubMed: 30926966]
- Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, ... Easton DF (2019). Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *The American Journal of Human Genetics*, 104(1), 21–34. 10.1016/J.AJHG.2018.11.002 [PubMed: 30554720]
- Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, ... Easton DF (2015). Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nature Genetics*, 47(4), 373–380. 10.1038/ng.3242 [PubMed: 25751625]
- Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, ... Easton DF (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature*, 551(7678), 92–94. 10.1038/nature24284 [PubMed: 29059683]
- Nathanson KN, Wooster R, & Weber BL (2001, 5). Breast cancer genetics: What we know and what we need. *Nature Medicine*. Nature Publishing Group. 10.1038/87876
- Shu X, Long J, Cai Q, Kweon SS, Choi JY, Kubo M, ... Zheng W (2020). Identification of novel breast cancer susceptibility loci in meta-analyses conducted among Asian and European descendants. *Nature Communications*, 11(1), 1–9. 10.1038/s41467-020-15046-w
- Speed D, & Balding DJ (2019). SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nature Genetics*, 51(2), 277–284. 10.1038/s41588-018-0279-5 [PubMed: 30510236]
- Speed D, Cai N, Johnson MR, Nejentsev S, & Balding DJ (2017). Reevaluation of SNP heritability in complex human traits. *Nature Genetics*, 49(7), 986–992. 10.1038/ng.3865 [PubMed: 28530675]
- Taliun D, Harris D, Kessler M, Carlson J, Szpiech Z, Torres R, ... Abecasis G (2019). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *BioRxiv*, 2, 563866. 10.1101/563866
- Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, & Jemal A (2015). Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*, 65(2), 87–108. 10.3322/caac.21262 [PubMed: 25651787]
- Vilhjálmsdóttir BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, ... Price AL (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *The American Journal of Human Genetics*, 97(4), 576–592. 10.1016/J.AJHG.2015.09.001 [PubMed: 26430803]
- Wang Y, Guo J, Ni G, Yang J, Visscher PM, & Yengo L (2020). Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nature Communications*, 11(1), 1–9. 10.1038/s41467-020-17719-y
- Wen W, Shu X, Guo X, Cai Q, Long J, Bolla MK, ... Zheng W (2016). Prediction of breast cancer risk based on common genetic variants in women of East Asian ancestry. *Breast Cancer Research*, 18(1), 124. 10.1186/s13058-016-0786-1 [PubMed: 27931260]
- Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, ... Carlson CS (2019). Genetic analyses of diverse populations improves discovery for complex traits. *Nature*, 1. 10.1038/s41586-019-1310-4

- Zheng W, Long J, Gao Y-T, Li C, Zheng Y, Xiang Y-B, ... Shu X-O (2009). Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nature Genetics*, 41(3), 324–328. 10.1038/ng.318 [PubMed: 19219042]
- Zheng W, Zhang B, Cai Q, Sung H, Michailidou K, Shi J, ... Kang D (2013). Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. *Human Molecular Genetics*, 22(12), 2539–2550. 10.1093/hmg/ddt089 [PubMed: 23535825]

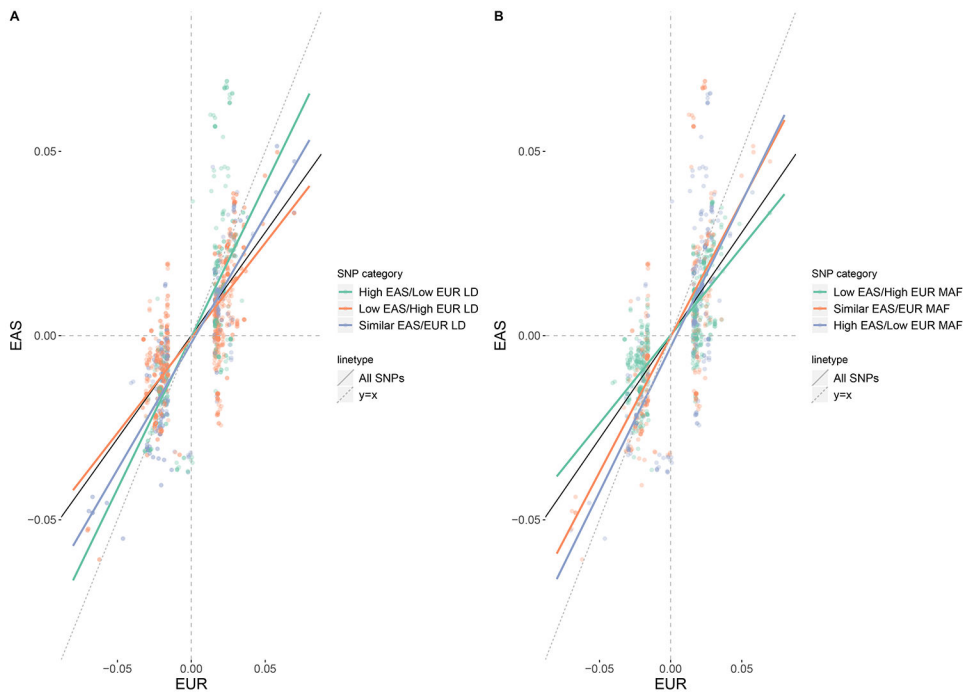




**Fig 1. The AUC of PRSs derived from the EAS/EUR GWASs alone and meta-analysis of both.** PRSs are derived from EAS, EUR GWASs and fixed effect meta-analysis of the EAS and EUR (denoted as META\_FE). PRSs are evaluated in an independent EAS validation dataset for predicting breast cancer risk. Each PRS was plotted against the (A)  $P$ -value threshold and (B) number of SNPs included. The corresponding numerical results were reported in S2 Table.

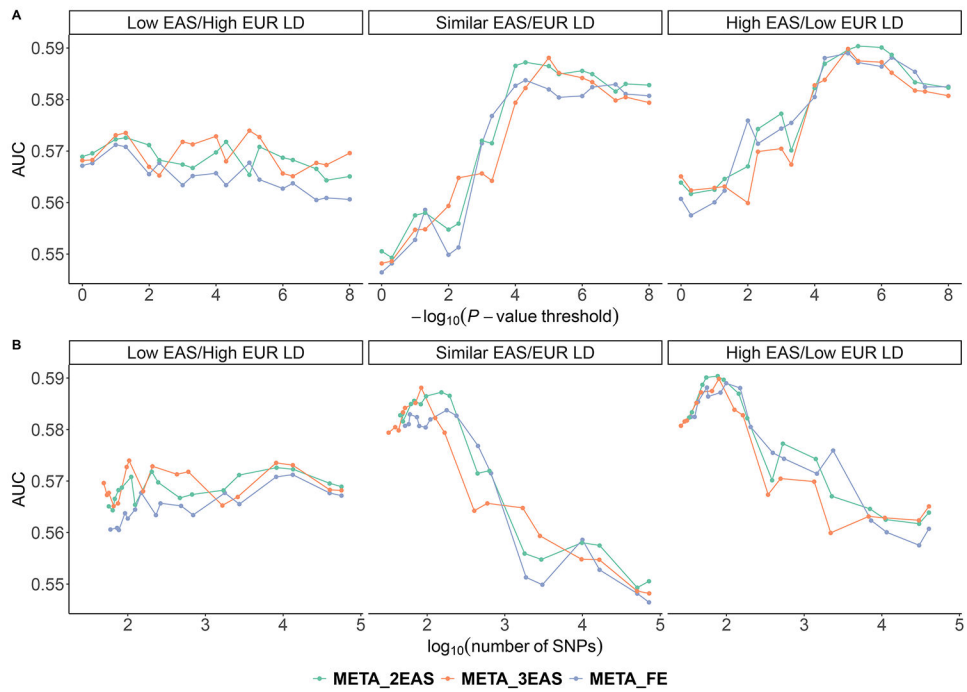


**Fig 2. The AUC of PRSs derived from rescaled meta-analysis of the EAS and EUR\_2015.** PRSs are evaluated in predicting breast cancer risk in the EAS validation dataset. META\_FE denotes the conventional fixed-effect meta-analysis; META\_2EAS denotes the rescaled meta-analysis that upweights the EAS effect size estimates by a factor of two; and META\_3EAS denotes the rescaled meta-analysis that upweights the EAS effect size estimates by a factor of three. Each PRS was plotted against the (A)  $P$ -value threshold and (B) number of SNPs. The corresponding numerical results were provided in S2 Table



**Fig 3. Standardized Z-scores of SNPs from EUR and EAS GWAS classified into different LD and MAF groups.**

SNPs with  $P$ -values  $< 5 \times 10^{-8}$  in either EUR\_2015 or the EAS GWAS were included. (A) Low EAS/High EUR LD: SNPs with  $I_{diff}$  less than its 33% quantile; Similar EAS/EUR LD: SNPs with  $I_{diff}$  between its 33% and 67% quantiles; High EAS/Low EUR LD: SNPs with  $I_{diff}$  larger than its 67% quantile. (B) Low EAS/high EUR MAF: SNPs with  $MAF_{diff,j}$  less than its 33% quantile, similar EAS/EUR MAF: SNPs with  $MAF_{diff,j}$  between its 33% and 67% quantiles, and high EAS/low EUR MAF: SNPs with  $MAF_{diff,j}$  larger than its 67% quantile. The black dashed line is the slope of the fitted line of standardized Z score in EAS over standardized Z score in EUR with all SNPs included; the red/blue/green lines are fitted lines of standardized Z score in EAS over standardized Z score in EUR for SNPs in corresponding LD group or MAF group.



**Fig 4.** The AUC of PRSs constructed with SNPs in different  $I_{diff}$  groups. All SNPs were classified into three groups: low EAS/high EUR LD score group with  $I_{diff}$  less than its 33% quantile, similar EAS/EUR LD score group with  $I_{diff}$  between its 33% and 67% quantiles, and high EAS/low LD score group with  $I_{diff}$  larger than its 67% quantile. META\_FE denotes the conventional fixed-effect meta-analysis; META\_2EAS denotes the rescaled meta-analysis that upweights the EAS effect size estimates by a factor of two; and META\_3EAS denotes the rescaled meta-analysis that upweights the EAS effect size estimates by a factor of three. Each PRS was plotted against the (A)  $P$ -value threshold and (B) number of SNPs.

**Table 1.**

Prediction AUC, Nagelkerke's  $R^2$  and liability adjusted  $R^2$  in the EAS validation set (adjusted for age and first 2 principal components of genotype)<sup>a</sup>.

Model <sup>b</sup>	<i>P</i> -threshold <sup>c</sup>	N_SNPs <sup>d</sup>	AUC <sup>e</sup>	AUC 95% CI <sup>f</sup>	Nagelkerke's $R^2$ <sup>g</sup>	Liability adjusted $R^2$ <sup>h</sup>	<i>P</i> -value for improvement over reduced model <sup>i</sup>
EAS	$5 \times 10^{-6}$	44	0.5782	0.5600-0.5963	0.047	0.0201	$9.64 \times 10^{-29}$
EUR	$5 \times 10^{-8}$	107	0.5809	0.5628-0.5989	0.0465	0.0198	$1.99 \times 10^{-28}$
META_FE	$1 \times 10^{-5}$	257	0.6008	0.5829-0.6188	0.0624	0.0268	$1.33 \times 10^{-38}$
META_2EAS	$5 \times 10^{-6}$	190	0.6049	0.587-0.6228	0.0655	0.0282	$1.30 \times 10^{-40}$
META_3EAS	$1 \times 10^{-5}$	193	0.6059	0.588-0.6238	0.0656	0.0283	$1.08 \times 10^{-40}$
ADD3	NA	265	0.6096	0.5917-0.6274	0.0689	0.0298	$7.70 \times 10^{-43}$

<sup>a</sup>We reported AUC, Nagelkerke's  $R^2$  and liability adjusted  $R^2$  for each of the PRSs in the EAS validation dataset

<sup>b</sup>Models that PRSs are based on: EAS: EAS GWAS derived PRS; EUR: EUR GWAS derived PRS; META\_FE: conventional fixed-effect meta-analysis; META\_2EAS: rescaled meta-analysis that up-weights the EAS summary statistics by a factor of two; META\_3EAS: rescaled meta-analysis that up-weights the EAS summary statistics by a factor of three; ADD3: summation of the three best PRSs within each LD category

<sup>c</sup>*P*-value cutoff for including SNPs in model

<sup>d</sup>Number of SNPs in model

<sup>e</sup>Area under ROC curve

<sup>f</sup>95% confidence interval of AUC

<sup>g</sup>Nagelkerke's  $R^2$  of a full model

<sup>h</sup>Liability adjusted Nagelkerke's  $R^2$  that accounted for case/control ratio

<sup>i</sup>*P*-value from likelihood ratio test compare full model (PRS, covariates including age and first 2 principal components of genotype) with a reduced model (a model with covariates age and first 2 principal components of genotype only)

**Table 2**

OR of developing breast cancer in the top 10% individuals with the highest PRSs versus the remaining 90% in the EAS validation dataset

Model <sup>a</sup>	OR	OR (95% CI) <sup>b</sup>	P-value
EAS	2.23	1.78 - 2.80	$2.33 \times 10^{-13}$
EUR	1.73	1.39 - 2.16	$3.93 \times 10^{-7}$
META_FE	2.50	1.99 - 3.14	$8.40 \times 10^{-17}$
META_2EAS	2.31	1.85 - 2.91	$1.86 \times 10^{-14}$
META_3EAS	2.43	1.94 - 3.06	$5.34 \times 10^{-16}$
ADD3	2.50	1.99 - 3.14	$8.40 \times 10^{-17}$

<sup>a</sup>META\_FE denotes the conventional fixed-effect meta-analysis; META\_2EAS denotes the rescaled meta-analysis that upweights the EAS effect size estimates by a factor of two; META\_3EAS denotes the rescaled meta-analysis that upweights the EAS effect size estimates by a factor of three; ADD3 denotes the *ldiff*-stratified PRS constructed by the summation of the three best PRSs within each *ldiff* group.

<sup>b</sup>confidence interval of OR.